

StatInf_Project1_Simulation_Exercise

Willianto Asalim

21/05/2020

Statistical Inference Proj 1: Simulation Exercise

Investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with `rexp(n, lambda)` where `lambda` is the rate parameter. The mean of exponential distribution is $1/\lambda$ and the standard deviation is also $1/\lambda$. Set $\lambda = 0.2$ for all of the simulations. Investigate the distribution of averages of 40 exponentials and perform a thousand simulations.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials using the following:

- Show the sample mean and compare it to the theoretical mean of the distribution.
- Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
- Show that the distribution is approximately normal.

```
##Setup variables provided by the instructions.
n <- 40 ##number of exponentials
lambda <- 0.2 ##universal lambda for all simulations
simNum <- 1000 ## number of simulations

set.seed(1234) ## set the seed to create reproducibility

##Sample Mean Simulation in matrix format
sim <- matrix(rexp(simNum * n, rate=lambda), simNum, n)
meanSim <- rowMeans(sim) ## Refer to Appendix for the sample mean
```

1. Show the sample mean and compare it to the theoretical mean of the distribution.

As stated in the instruction the expected mean, μ of a exponential distribution is $1/\lambda$. We are going to compare this with the sample mean.

The theoretical mean is

```
muTheory <- 1/lambda
muTheory
```

```
## [1] 5
```

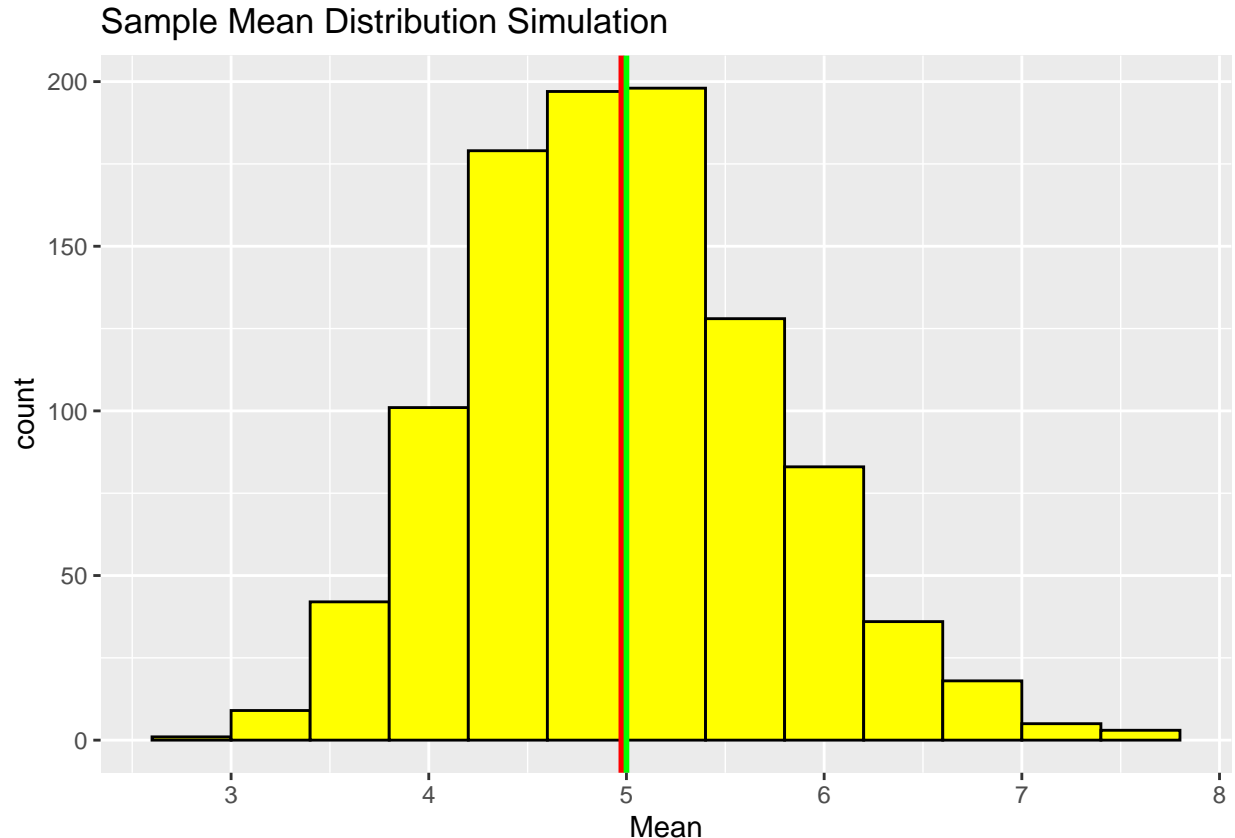
The sample mean is

```
muSim <-mean(meanSim)
muSim
```

```
## [1] 4.974239
```

Ans: The sample mean is 4.9742388 and the theoretical mean is 5 which prove that these values are very close. The plot below will give you a better picture of these two means.

```
ggplot(data.frame(meanSim), aes(x = meanSim)) +  
  geom_histogram(position="identity", color = "black", fill="yellow", binwidth=0.4) +  
  labs(title = "Sample Mean Distribution Simulation", x = "Mean") +  
  geom_vline(xintercept = muSim, size=1, colour="red") +  
  geom_vline(xintercept = muTheory, size=1, colour="green")
```



2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

To calculate the variance of Exponential Distribution, it is $\text{Standard_Deviation}^2/n$. Therefore the theoretical variance is $(1/\lambda)^2/n$

```
varTheory <- (1/lambda)^2/n  
varTheory
```

```
## [1] 0.625
```

The variance of the Sample mean is

```
varSample <- var(meanSim)  
varSample
```

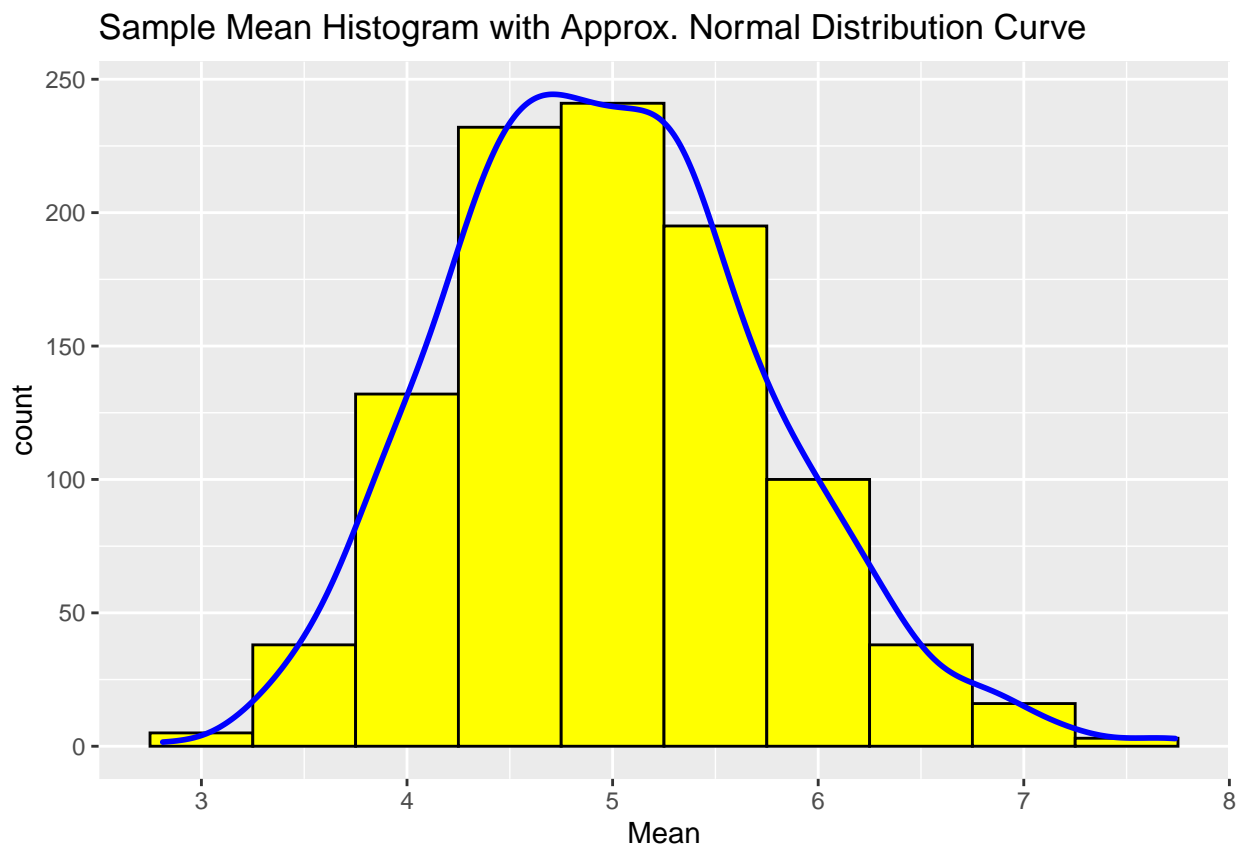
```
## [1] 0.5949702
```

Ans: From the calculation, we know that the sample distribution variance is 0.5949702 and the theoretical variance is 0.625 which is pretty close.

3. Show that the distribution is approximately normal.

The plot below shows that the the sample mean distribution is approximately normal by looking at the blue colour curve. The blue colour curve is approximately a normal distribution curve.

```
ggplot(data.frame(meanSim), aes(x = meanSim,)) +  
  geom_histogram(position="identity", color = "black", fill="yellow", binwidth=0.5) +  
  geom_density(aes(y=0.5*..count..),colour="blue", size=1) +  
  ##stat_function(fun = dnorm, colour = "green", geom = "point", args = list(mean = muTheory, sd=sqrt(v  
  ##scale_y_continuous(breaks=c()) +  
  ##scale_x_continuous(breaks=c(3, 4, 5, 6, 7, 8), limits=c(3, 8)) +  
  ##geom_vline(xintercept = muSim, size=1, colour="red") +  
  ##geom_vline(xintercept = muTheory, size=1, colour="green") +  
  labs(title = "Sample Mean Histogram with Approx. Normal Distribution Curve", x = "Mean")
```



Ans: The central limit theorem states that if you have a population with mean and standard deviation and take sufficiently large random samples from the population with replacement, then the distribution of the sample means will be approximately normally distributed. The plot above proved the Central Limit Theorem thus we can use the normal probability model

with confidence to quantify uncertainty when making inferences about a population mean based on the sample mean.

4. Appendix

First 6 lines of the mean of simulation 4.6025099, 6.0177897, 5.4636863, 4.1767548, 7.1446716, 4.4275673

The platform specification used:

Spec	Description
OS	Windows 10 Pro - 64 bit
CPU	AMD Ryzen 5 - 3400G
RAM	16GB DDR4 3000MHz
Storage	500GB SSD - M.2 NVMe (PCIe)